

Brief Guide to Data Tabulations

This document serves as a brief guide for understanding the data tables accompanying the report the Board of Governors prepared in response to the Congressional directive in the Gramm-Leach-Bliley Act of 1999. The tables include tabulations of responses to the survey.

The data set has two important features. First, because of the size and complexity of the survey, the names used to identify variables in the data set are not exactly the same as those of the questionnaire. This difference is due to the need to encode the column location, question subsection, and response category into variable names to ensure a unique identifier for each data item.

Second, to standardize the codes used to indicate survey responses, individual questions are often represented as multiple variables in the data set. Many questions in the survey offer respondents several response options or have several parts to them. Each of these options or parts is considered a *sub-part* of the question and is identified by a separate variable in the data set. Thus, a question that offers respondents five possible answers is translated into 5 distinct variables in the data set. Responses to these questions are coded either as a “1” for yes or “2” for no, where “1” indicates that the respondent chose the option. For questions that a respondent was restricted to choose only one option, only one of the options will be coded as “1”; all other options will be coded as “2.” For questions that a respondent was permitted to choose more than one option, several options may be coded as “1.”

1. Translating Survey Questions into Data Set Variables

With a few exceptions, the mapping from the questionnaire variable names to the data set variable names is straightforward. This section explains the basic mapping scheme used and points out the

few exceptions where the variable name cannot be directly deduced from the questionnaire.

Each variable name is seven characters long. The first two characters indicate the part and section of the questionnaire from which the variable is derived. The questionnaire is divided into two parts, A and B. Part A has four sections, so variable names in these sections begin with A1, A2, A3, A4. For example, all questions in the small business lending section, which is the third section of Part A of the survey, begin with "A3." Although Part B has only one section, two tabulations are provided. The first tabulation includes summary information on all CRA special lending programs that each banking institution operates (column F in the survey). Variable names in this section begin with B1. The second tabulation includes the information from each of the 341 programs for which detailed information was provided. Variable names in this section begin with B2.

The third and fourth characters of the variable name indicate the question number within a particular section of the survey. Each question number is required to have 2 digits. For example, variables associated with the fourth question have "04" as the third and fourth characters of their variable name.

The fifth character is used to identify any sub-parts question might have. For example, question 22 of section A1, which has four sub-parts, is translated into 4 variables in the data set that begin with A122A, A122B, A122C, and A122D. If there are no sub-parts to the question, a zero is used as a placeholder. The exception to preserving this label is where a question has only a single subpart. In these cases, the letter designating the sub-question is superfluous and a '0' is used instead. There are a handful of cases where the sixth character is also used to indicate the question's sub-part. (For example, question 12 of section A4 has sub-parts A1, A2, B1, B2, C1, C2, and D.)

Except for the few cases mentioned above, the sixth character of the variable name is “0.”

The seventh character contains the corresponding column letter for the response--either “A” or “B”. If there is only one column for a question, the seventh character of the variable name contains a '0'.

Some examples of the naming translation are presented in the box “Examples of the translation from survey questions to data set variables.”

Examples of the translation from survey questions to data set variables

1. The variable corresponding to the first question in section 1 of part A is named A101000.
2. Question five of the same section has 4 associated variables:
column A: A105A0A, A105B0A
column B: A105A0B, A105B0B
3. In a similar manner, question 9 in that section has ten corresponding variables (5 for each column):
A109A0A, A109B0A, A109C0A, A109D0A, A109E0A
A109A0B, A109B0B, A109C0B, A109D0B, A109E0B
4. Question 12 in section 4 of part A is a slightly more complicated case. There are eight variables associated with this question:
A412A10, A412A20, A412B10, A412B20, A412C10,
A412C20, A412D00

Four other variables are included in the data tables. SIZE indicates the breakdown of responding institutions by the three asset-size categories used in the report (“1” is the smallest asset-size category, “3” is the largest asset-size category). CENSUS indicates the Census region in which the banking institution’s headquarters is located (“1” is Northeast, “2” is Midwest, “3” is South, and “4” is West). THRIFT indicates whether a respondent was a savings associations (“1”

is yes, “2” is no). In section B1, B101PRO indicates the number of CRA special lending programs each banking institution reported that they were involved in. This variable is top-coded at 5.

2. Variable codes

There are two standard entries for all variables in the data tables. Some questions ask for quantities while others ask respondents to choose between several options. For questions seeking quantities, variables indicate the mean, median, and standard deviation for all responses, as well as the number of responses that are 0. For questions seeking choices between options, the options have been numbered as 1 through the number of options for the question. The most common type of question is a question which asks respondents to indicate yes or no. For these variables, yes is coded as “1” and no is coded as “2.” For questions asking respondents to rank activities (for example, “higher,” “about the same,” “lower”), the first option is typically coded as “1” and numbers are assigned as options appear. An exception to this convention occurs for follow-up questions to qualitative question that appear in the final portions of sections A1 through A3. For these questions, which have 3 columns for each sub-part, “Higher for CRA” is coded as “3,” “About the same” is coded as “2,” and “Lower for CRA” is coded as “1.”

All respondents did not provide answers to every question. The data indicate two classes of non-response. First, a set of questions or an entire section might not be applicable for a respondent. Examples of this would be if a particular banking institution extended no loans in a given product category or did not purchase any loans for in a given product category. These are coded as “A” in the data tables. Second, a respondent may not have enough information to answer a particular question, may have refused to answer particular questions, or may have

overlooked a question. These types of non-response are coded as “U” in the data tables.

Tabulations for each response appear in columns titled “COUNTX,” where X represents the actual response. For questions that seek quantities, the data tables also indicate the number of responses that were used in calculating the quantities. This number appears in the column titled “COUNTN” in the tables.